

Robust abandoned object detection integrating wide area visual surveillance and social context

Article

Accepted Version

Ferryman, J., Hogg, D., Sochman, J., Behera, A., Rodriguez-Serrano, J. A., Worgan, S., Li, L., Leung, V., Evans, M., Cornic, P., Herbin, S., Schlenger, S. and Dose, M. (2013) Robust abandoned object detection integrating wide area visual surveillance and social context. *Pattern Recognition Letters*, 34 (7). pp. 789-798. ISSN 0167-8655 doi: <https://doi.org/10.1016/j.patrec.2013.01.018> Available at <https://centaur.reading.ac.uk/31241/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <http://www.sciencedirect.com/science/article/pii/S0167865513000226>

To link to this article DOI: <http://dx.doi.org/10.1016/j.patrec.2013.01.018>

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Robust abandoned object detection integrating wide area visual surveillance and social context

James Ferryman^{a,*}, David Hogg^b, Jan Sochman^b, Ardhendu Behera^b, José
A. Rodriguez-Serrano^c, Simon Worgan^d, Longzhen Li^a, Valerie Leung^f,
Murray Evans^a, Philippe Cornic^e, Stephane Herbin^e, Stefan Schlenger^g,
Michael Dose^g,

^a*Computational Vision Group, School of Systems Engineering, University of Reading,
RG6 6AY, UK*

^b*School of Computing, University of Leeds, UK*

^c*Xerox Research Centre Europe, 6 Chemin de Maupertuis, 38240 Meylan, France*

^d*formerly University of Leeds*

^e*Department of Information Processing and Modelling, ONERA, BP 80100, 91123
Palaiseau Cedex, France*

^f*MathWorks, Les Montalets, 2 rue de Paris, 92190 Meudon, France*

^g*L-1 Identity Solutions, Universitaetsstr.160, 44801 Bochum, Germany*

Abstract

This paper presents a video surveillance framework that robustly and efficiently detects abandoned objects in surveillance scenes. The framework is based on a novel threat assessment algorithm which combines the concept of ownership with automatic understanding of social relations in order to infer abandonment of objects. Implementation is achieved through development of a logic-based inference engine based on Prolog. Threat detection performance is conducted by testing against a range of datasets describing realistic situations. The proposed system represents the approach employed in the EU SUBITO project (Surveillance of Unattended Baggage and the Identification and Tracking of the Owner).

Keywords:

^{*}This document is a collaborative effort.
Preprint submitted to Pattern Recognition Letters
Corresponding author; email: james@computer.org

January 10, 2013

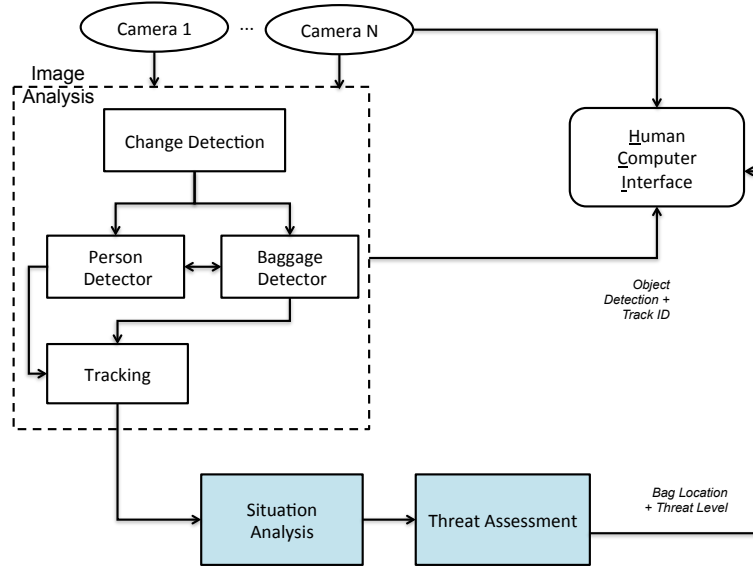


Figure 1: General framework of the automated threat detection system

28 wide area video surveillance, behaviour analysis, abandoned objects

29 1. Introduction

30 In recent years there have been a number of incidents where terror organ-
 31 isations have planted explosive devices in ordinary baggage to cause immense
 32 disruption in mass transportation networks and other areas of critical infras-
 33 tructure. Due to the potentially devastating consequences of such terrorist
 34 activity, the monitoring and surveillance of unattended baggage has become
 35 a priority for the security operators of mass transportation networks and
 36 other critical infrastructure. The overriding goal is to minimise the number
 37 of false alarms. Towards this goal, the main contribution of this work is
 38 the development and evaluation of behaviour analysis methodology permit-

39 ting robust identification of a baggage-owner while minimising false positives.
 40 The approach taken advances the state of the art in abandoned bag detec-
 41 tion by introducing the concept of ownership and combines it with automatic
 42 understanding of social groups to infer abandonment. To achieve the goal, a
 43 framework (see Figure 1) has been developed consisting of a complete four-
 44 fold process, detection - tracking - situation analysis - threat assessment.
 45 This paper is divided as follows. Firstly, in Section 2 related research is de-
 46 tailed, followed in Sections 3-5 by descriptions of the system components. In
 47 Section 6 the datasets used and results of experiments are presented before
 48 concluding in Section 7 with conclusions and recommendations for future
 49 research.

50 **2. Related Work**

51 There exists a significant body of academic research addressing the task
 52 of robustly identifying abandoned baggage in public spaces. Most authors
 53 treat detection of abandoned (or left) objects, especially luggage, as the task
 54 of static object detection, with (Birch et al., 2011; Tian et al., 2010) or with-
 55 out (e.g. (Evangelio and Sikora, 2011; Porikli et al., 2008)) the application
 56 of tracking. Tian et al. (2010) present a framework to detect abandoned
 57 and removed scene objects based on background subtraction and foreground
 58 analysis, combined with tracking output to reduce false positives. Birch et al.
 59 (2011) employ motion segmentation based on a GMM with fast learning and
 60 a Motion History Image (MHI). For tracking of stationary objects, the edge
 61 map (3x3 Sobel filter) for each pixel is computed and matched) by correla-
 62 tion of edge directions. A comparative evaluation of stationary foreground

63 detection algorithms based on background subtraction is given in Bayona et
64 al. (2009).

65 There has been some attempt at human activity recognition and associ-
66 ation to scene objects. In Lu et al. (2007) moving objects are tracked using
67 shape and colour features and Kalman-based filtering, and classified using
68 eigen features and Support Vector Machine. A package is defined as a non-
69 human object and package ownership analysis performed using HMM-based
70 human activity recognition.

71 *2.1. Dataset Based Challenges*

72 The most widely used datasets with which to evaluate approaches to
73 abandoned bag detection have been from (PETS2007; PETS2006) and from
74 the UK Home Office i-LIDS (2007). The dataset provided for the PETS2006
75 challenge consists of 7 multi-camera scenarios involving an increasing num-
76 ber of people and passers-by. Most of the submissions to PETS2006 were
77 based on background subtraction combined with a blob tracker (Auvinet et
78 al., 2006; Guler and Farrow, 2006; Krahnstoever et al., 2006; Li et al., 2006;
79 Martínez-del-Rincín et al., 2006; Smith et al., 2006), with the exception of
80 Lv et al. (2006) who rely on a more realistic human model by incorporating
81 a human detector. Most often, when an object is not moving and its size
82 is beneath a given threshold, it is assumed to be a standing bag. Smith
83 et al. (2006) propose a probabilistic approach in which people and bags are
84 classified based on the immediate history of their size and velocity. Another
85 approach from PETS2006 is to use a slow-decay background model to de-
86 tect stationary objects (Guler and Farrow, 2006). To be able to apply the
87 PETS2006 rules for abandoned baggage (the owner is further than a metres

88 for more than b seconds), the owner is usually defined as the nearest tracked
89 object when the standing bag appears (Krahnstoevers et al., 2006; Lv et al.,
90 2006) or by examining blob splits during tracking (Auvinet et al., 2006; Guler
91 and Farrow, 2006; Smith et al., 2006). When a standing bag and its owner
92 are identified, it is straightforward to apply the PETS2006 abandoned-bag
93 rules. The simplicity of the scenarios allows very limited situation aware-
94 ness and was designed mainly to test if the low level processing stages are
95 sufficient to cope with real-world scenarios.

96 The PETS2007 challenge focusses on two additional scenarios: theft and
97 loitering. The videos are much more challenging from the tracking point of
98 view as the scenes are more crowded. There are 8 scenarios, each viewed from
99 4 cameras. Two submissions to the challenge go beyond classical approaches
100 to blob tracking and split-track analysis (such as (Arsic et al., 2007; Dalley et
101 al., 2007)) and slowly/quickly adapting background models (such as Porikli
102 and Yin (2007)). Firstly, Ribeiro et al. (2007) use a Temporal-JointBoost
103 algorithm for each blob being tracked to classify it into a person-walking, not
104 moving, a person picking-up/leaving a bag, or an abandoned bag. The basic
105 idea is to incorporate temporal features (optical flow, motion energy) into the
106 classification process over some temporal window. Secondly, Ardo and As-
107 trom (2007) use an HMM to improve the temporal consistency of the tracking
108 and show how to use an HMM efficiently in this setting. These approaches
109 demonstrate the potential advantages of considering a longer temporal win-
110 dow for activity analysis. Nevertheless, the situation awareness in the PETS
111 2007 challenge is again very simple - reduced to comparing the distance of a
112 bag to its owner (abandoned bag, theft) or measuring the time for which a

113 person stays in the scene (loitering).

114 The UK Home Office have developed an image library (i-LIDS, 2007) to
115 help researchers and designers to evaluate video based detection systems to
116 meet Government requirements. The i-LIDS library includes an abandoned
117 luggage dataset including several challenges of single instances of left lug-
118 gage on a metro platform in the presence of passing passengers and trains.
119 While the dataset is useful for evaluating detection algorithms it remains lim-
120 ited because it is monocular and also does not contain examples of specific
121 behavioural interactions.

122 *2.2. Limitations of Existing Approaches*

123 It is clear that a global analysis of the situation rather than just ex-
124 amining each agent’s behaviour independently, would be beneficial in many
125 situations. The motivation for this is illustrated by a scenario similar to
126 that of (PETS2007) where a family or a group of friends comes together and
127 one of them leaves his/her bag with the others. Any threat detection system
128 treating the individuals independently would inevitably report an abandoned
129 bag, as the criteria specified in (PETS2006) that the bag is abandoned if the
130 owner is further than a metres for more than b seconds, is fulfilled. For treat-
131 ing these more complex scenarios, the approaches described above may be
132 insufficient and it may be necessary to derive a more complete activity anal-
133 ysis. A significant corpus of the computer vision and artificial intelligence
134 literature attacks the problem of understanding activities from visual input.
135 While logic and grammar-based representations, with or without combina-
136 tion with statistical approaches, (Hongeng et al., 2004; Ivanov and Bobick,
137 2000; Joo and Chellappa, 2006; Shet et al., 2005) organise knowledge in a

flexible, powerful and clean way, one drawback of these approaches is that they are unable to propagate the uncertainty in the primitive detections. Hidden Markov Models (Brand et al. (1997)) and other flavours of dynamic Bayesian network provide a powerful generalisation of stochastic finite state automata to deal with such uncertainty. Another related approach is the so-called propagation network (Shi et al., 2004). In recent work, Damen and Hogg (Damen, 2012) first specify activities using a multiset attribute grammar and then convert it to an equivalent Bayesian network. A more general tool which converts first-order logic predicates into an equivalent Bayesian network is the framework of Markov logic networks (Richardson and Domingos, 2006), which have also been applied to activity analysis (Tran, 2008). An entirely different approach is to detect events from image pixels directly rather than by reasoning about the interactions between specific agents, for instance (Li, 2008; Wang, 2009). Whilst these approaches are easily configured to output whether an activity is normal or abnormal, they lack the explanatory power of grammar and logic-based methods (i.e. why it is abnormal).

None of the approaches described in the literature, however, have combined the concept of ownership with recognition of social groups, to reduce the number of false positives in detection of abandoned objects.

3. Object Detection and Tracking

The framework, shown in Figure 1, supports application of a range of object detectors and trackers including the POM person detection method of Berclaz et al. (2009) and tracking-by-detection of Breitenstein et al. (2011),

162 both of which operate at low frame rates (2-4fps) or offline. While detection
 163 and tracking is not the main contribution of this paper, brief descriptions are
 164 given to methods which have been developed to permit the overall framework
 165 to operate online and with multiple cameras.

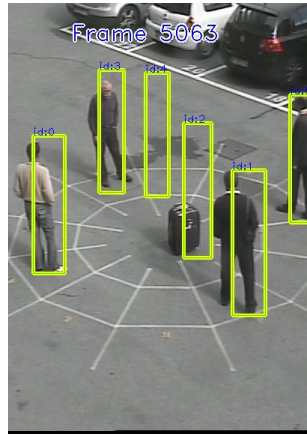
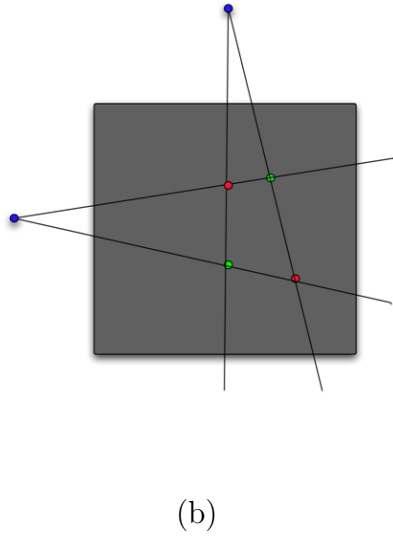
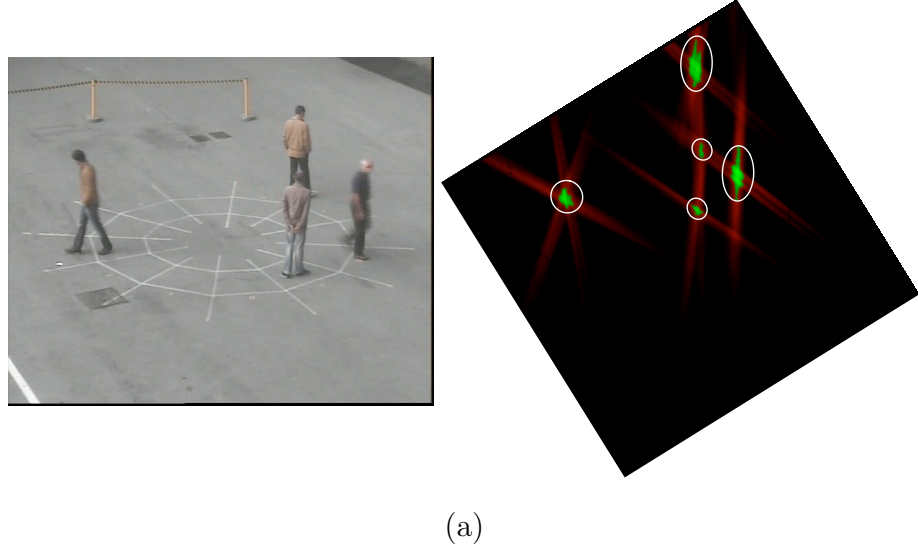
166 *3.1. Baggage Detection*

167 Baggage hypothesis generation is based on static change detection using
 168 the dual background approach of Porikli et al. (2008) adapted to use the
 169 efficient implementation of the Gaussian Mixture Model in Zivkovic (2004).
 170 Bag verification consists of application of a combination of filters including
 171 both 2D and 3D geometric filters and foreground/background similarity filter,
 172 and temporal filtering to check for persistence of the static regions.

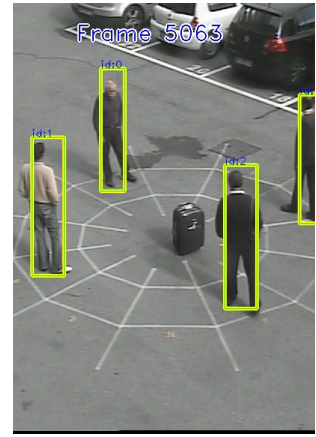
173 *3.2. Person Detection*

174 Person detection is based on the homography based multi-camera ap-
 175 proach of Yildiz and Akgul (2010), extended with a novel approach for ghost
 176 suppression. First, a synergy map, the result of projecting detected fore-
 177 ground from each camera view to a single plane, is created, as shown in
 178 Figure 2. In practice, the reverse process is used with sampled cells on the
 179 synergy map, each corresponding to a vertical cuboid in space of fixed person
 180 height, back-projected to the bounded rectangles in the original images. The
 181 process is applied for an image resolution-limited "infinite" number of planes
 182 in a very efficient and fully real-time manner without hardware acceleration.

183 For a given location (x, y) in the Synergy map (which corresponds to a
 184 small rectangular region on the ground plane), the value $S(x, y)$ accumulating
 185 the evidence of a person's presence can be calculated as:



(c)



(d)

Figure 2: Synergy map: (a). Detection of all pedestrians requires a threshold on synergy map to be set to value that permits ghost detection to pass thorough. (b). Ghost positions (red) can be predicted if correct positions (green) are known or can be estimated. (c-d). Bounding boxes resulting from detections without (c) and with (d) ghost prediction and suppression, for the same frame of video.

$$S(x, y) = \frac{1}{|I|} \sum_{i \in I} \frac{\sum_{u=u_0}^{u_1} \sum_{v=v_0}^{v_1} p(u, v, i)}{A(Z(x, y, i))} \quad (1)$$

186 where I is the set of images into which the cuboid can be visibly projected,
 187 $Z(x, y, i) = \{(u_0, v_0), (u_1, v_1)\}$ is the bounding box projection of the cuboid
 188 corresponding to a specific synergy map pixel (x, y) into image i as defined
 189 by two extreme corner points. $A(s)$ is a function to calculate the area of any
 190 shape s , and

$$p(u, v) = \begin{cases} 1, & \text{if } I(u, v) \text{ is foreground} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

191 Candidate objects are represented by peaks in the synergy map, obtained
 192 via thresholding. Ghost detections can occur where lines from different cam-
 193 eras to different objects intersect. To prevent ghosts becoming new tracking
 194 targets, a suppression map is generated in the regions of high ghost probabil-
 195 ity and subtracted from the synergy map. Frame-to-frame tracking of peaks
 196 further reinforces probable objects' location.

197 3.3. Tracking

198 A multi hypothesis tracker is used Blackman (2004) modified for appli-
 199 cation to tracking of extended objects. First, to handle short-term occlusions
 200 and the merging of measurements from different persons in the detection pro-
 201 cess, measurement-sharing between track hypotheses is allowed. This concept
 202 is illustrated in Figure 3 (Top). Secondly, the measurement-to-track associa-
 203 tion cost is modified to allow image features, specifically two hue-saturation
 204 histograms corresponding to the top and bottom halves of a person, to be
 205 used in addition to a simple Brownian motion model. Each model is updated

206 using the Exponentially Weighted Moving Average (EWMA). The associa-
207 tion score between a predicted state and a measurement is a product of the
208 normalised histogram intersection distance between their histograms and the
209 normalised Euclidean distance between their positions in 3D.

210 To overcome track fragmentations caused by long-term or complex pat-
211 terns of interaction between people, long term tracking based on tracklet
212 association is used. The approach is based on a Markov Logic Network
213 (MLN) (Leung and Herbin (2011)) where the notion of a group to account
214 for generic interaction between people is introduced. The scores for possible
215 associations are calculated using both spatial-temporal constraints and ap-
216 pearance information. Associations are not only considered for tracklets that
217 can be directly joined together; but are extended to tracklets separated by
218 a group in space and time. It therefore handles the formation and splitting
219 of groups, reducing track fragmentations and allowing longer tracks to be
220 formed. Examples of the tracklet association rules are shown in Figure 3
221 (Middle) and example final tracking output in Figure 3 (Bottom).

222 4. Situation Analysis

223 Situation analysis is an intermediate step towards threat assessment and
224 is defined as the description of the relationships between people and bags
225 that can be inferred from the behaviour of the participating agents. This
226 contribution focusses on two kinds of relationship: who owns each bag, and
227 who knows who. The analysis takes object tracks and class information as
228 input and describes the state of the world (i.e. the scene) in terms of the
229 observed agents and their behaviour. The following stage (threat assessment)

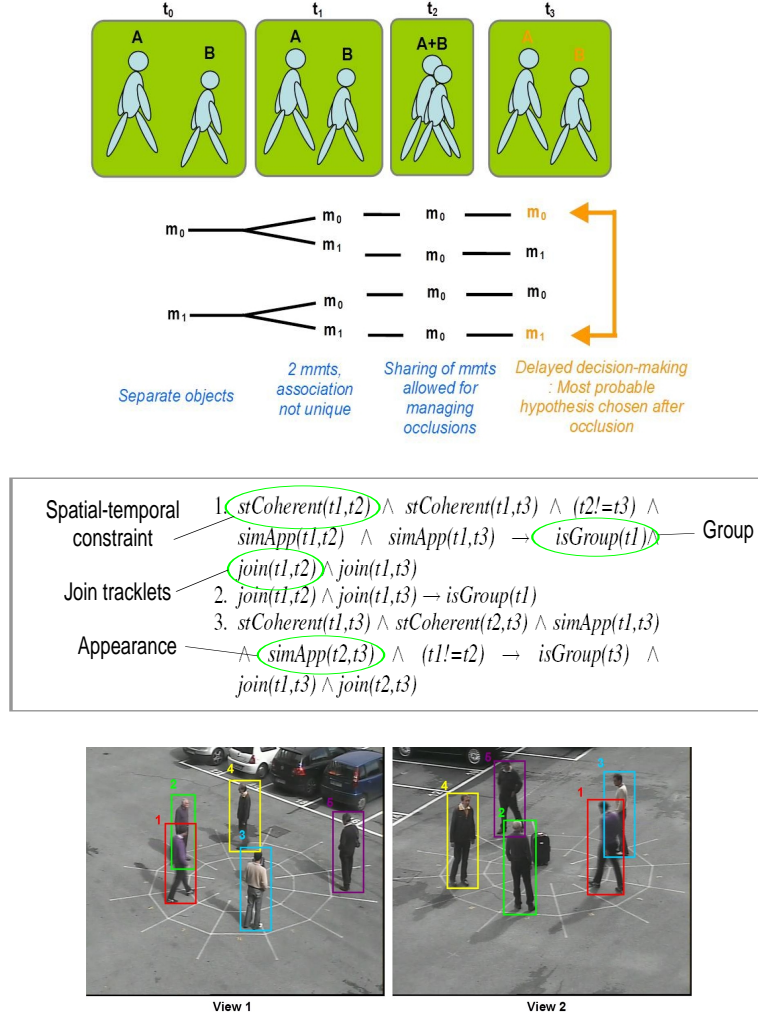


Figure 3: Tracking processes. Top: Illustrating how measurement-sharing in video-MHT overcomes short-term occlusions. Middle: Examples of tracklet association rules used in the MLN formalism. Spatial-temporal coherence and appearance information are used as inputs. The inference of groups and the joining of tracklets are two of the outputs. Bottom: Example tracking output for two cameras showing objects IDs.

determines whether the state of the world constitutes a possible threat (i.e. there is a truly abandoned bag.) The main contribution is the combination of the automatic understanding of social relationships with the concept of ownership to reduce the number of false alarms.

4.1. Bag Ownership

For the reported experiments in this paper, a bag is detected when it appears stationary in the scene, having been placed there by a person. At this stage, detection of a bag as it is carried into or out of the scene has not been incorporated. The ownership of each bag is inferred by simply looking for a person in the proximity of the bag over a fixed time interval prior to its appearance. The person is also required to be stationary at the time the bag-drop is hypothesised to occur. Specifically, in the experiments reported here, any person is assumed to be an owner if they are temporarily stationary within one metre of the bag at any point within one second prior to its appearance. Note that multiple possible owners are allowed, not because this is expected to be the case in reality but in order to reduce false alarms through taking both hypotheses through into the threat assessment.

4.2. Inference of Social Relations

Social groups are a very common phenomena in human crowds, with empirical studies suggesting that about 74% of people come in a group to a social event (Aveni (1977)) and about 50-70% (depending on the environment) are in a group during casual walking (Rudloff et al. (2011)). Despite this high percentage, the prevailing crowd behaviour models in todays simulation tools (Challenger et al. (2009)), computer graphics applications (Reynolds (1987))

254 and in particular in activity recognition and computer vision (PETS2006)
255 are based on modelling each individual independently. An online algorithm
256 has been developed for automatic detection of social groups within crowds,
257 based on the analysis of the way the social relations influence the walking
258 behaviour of the group members.

259 The method is based on the Social Force Model (SFM) (Helbing and
260 Molnar, 1995; Moussaid et al., 2010) widely used in the crowd simulation
261 community. In this, each individuals’ movement is influenced by notional
262 forces operating between individuals. Depending on whether two individ-
263 uals (a) know each other or (b) do not know each other, the Social Force
264 Model produces different sets of trajectories for these individuals. Until re-
265 cently, these attempts were based on human designed forces without proper
266 evaluation. Only recently, the model has been calibrated on real-world video
267 sequences resulting in a model that realistically predicts avoidance behaviour
268 of a walking group (Moussaid et al., 2009; Singh et al., 2009) and later in
269 a model with all its parameters, including group behaviour, estimated from
270 real data (Moussaid et al. (2010)).

271 The method employed in this work solves the inverse problem: knowing
272 the trajectories, what are the social forces, and thus the relations, that caused
273 that behaviour. The method is used in the framework to infer the social
274 relations between the individuals in a scene and thereby to inform threat
275 assessment as explained in Section 5.

276 The authors are aware of only two approaches aiming explicitly at social
277 group inference (Ge et al., 2009; Jacques et al., 2007) and one paper using
278 social groups to improve tracking (Pellegrini et al. (2010)). In Jacques et al.

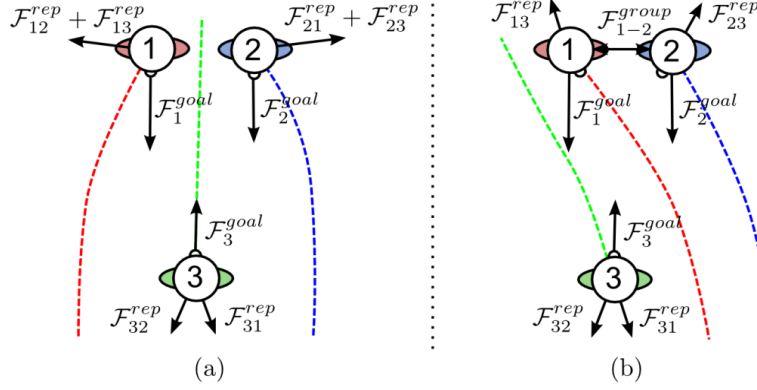


Figure 4: Depending on whether the individuals 1 and 2 (a) do not know each other or (b) know each other, the Social Force Model produces different sets of trajectories combining together repulsive (\mathcal{F}^{rep}), goal directed (\mathcal{F}^{goal}), and group (\mathcal{F}^{group}) forces influencing the individuals.

279 (2007) the groups are detected when two individuals keep close enough for
 280 a significant fraction of time over a given period. Experiments undertaken
 281 by the authors have shown that such simple measures are not sufficient for
 282 reliable group inference in complex scenes. In the proposed approach the
 283 calibrated SFM instead is relied upon. Similar measurements were used in
 284 Pellegrini et al. (2010) to improve tracking by jointly tracking and inferring
 285 the social groups.

286 Also based on distance, but including the difference in velocity as well
 287 as position, the method proposed in Ge et al. (2009) applies clustering to
 288 the (complete) person trajectories. The merging criterion takes into account
 289 the fraction of time in which the individuals are seen close to each other and
 290 allows the addition of a person to the group only if they have been close to
 291 at least half of its members. Figure 4 illustrates the Social Force Model. Full

292 details of the approach are given in Sochman and Hogg (2011).

293 5. Threat Assessment

294 The threat assessment stage determines whether the inferred situation
295 constitutes a threat, utilising the inferred knowledge of ownership and social
296 relations described in Section 4. The mechanism adopted is sufficiently gen-
297 eral to accommodate external information (e.g. the state of alert, time of
298 day) alongside information on the observed scene in determining whether or
299 not to raise an alarm.

300 Three increasingly sophisticated definitions are considered for what con-
301 stitutes an abandoned bag. The first adopts the simple *baseline* definition
302 that defined the PETS2006 challenge. In this, a threat (i.e. abandonment)
303 is defined as follows:

- 304 • *Bag unattended if no person within 2 metres*
- 305 • *Bag abandoned if unattended for 30 seconds*

306 Here, the notions of ownership and social relationships are not used.

307 The second definition (*owner*) includes the notion of ownership (Sec-
308 tion 4.1) and is defined as follows:

- 309 • *Bag unattended if owner is not within 2 metres*
- 310 • *Bag unattended if there is no assigned owner and if no person within 2*
311 *metres*
- 312 • *Bag abandoned if unattended for 30 seconds*

313 When there is no assigned owner, this is equivalent to the baseline def-
314 nition, but where one or more possible owners have have been assigned, the
315 condition for an alarm to be raised is less stringent since the behaviours of
316 non-owners within the scene is ignored (unless there is no assigned owner).

317 The third definition (*owner+group*) includes both the notions of owner-
318 ship (Section 4.1) and social relationships (Section 4.2). In this, a threat is
319 defined as follows:

- 320 • *Bag unattended if owner or someone in the same social group as owner*
321 *is not within 2 metres*
- 322 • *Bag unattended if there is no assigned owner and if no person within 2*
323 *metres*
- 324 • *Bag abandoned if unattended for 30 seconds*

325 This relaxes the *owner* definition in the direction of the *baseline* definition,
326 since now the circle of people attending to a bag is widened to include people
327 in the same group as the possible owner(s). The likelihood of raising an
328 alarm is therefore reduced.

329 5.1. Implementation

330 The aim in threat assessment is to make it straightforward to encode
331 the evolving state of the world and explore different behavioural patterns
332 that constitute a potential threat. To achieve this, a simple logic-based
333 inference system (Prolog) is adopted in which the current state of the world
334 is represented by a set of facts and the behavioural patterns that constitute
335 potential threats are encoded as rules.

336 The elements of this logic-based approach are:

- 337 • Facts (logical atoms), which are employed to describe situations. A fact
338 is of the form $R(A,B,\dots)$, where R indicates a type of relation between
339 the elements inside the brackets.
- 340 • Rules, which are employed to infer new facts from existing ones.

341 Given these elements, the threat assessment proceeds in two steps:

- 342 1. Tracking and detection data are converted into a set of facts;
- 343 2. A set of pre-defined rules is invoked to infer additional facts.

344 The position of an object in each frame is represented by a unique ID for
345 the object, it's class (person or bag), it's x,y position on the ground-plane
346 and the frame number:

347 $track(id, class, x, y, frame)$.

348 The social relationships between individuals are represented by a single
349 predicate that records a unique group ID for each person. This partitions the
350 set of people into social groups. Any person not assigned to a social group
351 is assumed to be outside any group. This is represented simply by facts of
352 the form:

353 $group(id, group_id)$.

354 For convenience, a 'class' predicate is used (as in $class(id, person)$.) to
355 record the class of each object independently of the 'track' facts.

356 The ownership of bags is inferred next by a set of Prolog rules that embody
357 the criteria described in Section 4.1. The result is a new set of facts, each
358 representing the ownership of a bag (b) by a person (p):

359 $owner(p, b).$

360 Finally, the alarm condition for the chosen threat definition is posed as a
361 Prolog query. As part of this, for the baseline definition, the condition that
362 a bag is attended translates into the rule:

363 $attended(B, T) :- class(P, person), nearby(P, T, B, T, 2).$

364 Here the rule states that a bag is attended at time T (shown on the left
365 of the ‘:-’) if it is owned by someone (call them P), and the position of P at
366 time T is within 2 metres (i.e. nearby) of the position of B at time T (shown
367 on the right of the ‘:-’). Upper case arguments are used to signify that these
368 are variables.

369 The equivalent set of rules for the *owner+group* definition, incorporating
370 the notions of ownership and social relationships, is as follows:

371 $attended(B, T) :- owner(P, B), nearby(P, T, B, T, 2), !.$

372 $attended(B, T) :-$

373 $\quad \backslash +owner(-, B), track(P, person, -, -, T), nearby(P, T, B, T, 2).$

374 $attended(B, T) :- owner(P, B), knows(P, Q), nearby(Q, T, B, T, 2), !.$

375 $knows(P, Q) :- group(P, G), group(Q, G).$

376 The first rule states that a bag B is attended at time T if there is an
377 owner P for the bag and this person is nearby. The second rule invokes the
378 baseline notion of being attended when there is no owner - the meaning of
379 ‘\+’ before the owner predicate means that this isn’t present in the database.
380 The third rule states that a bag is attended (at time T) if there is a second
381 person Q who is nearby the bag and P and Q know one another. The fourth
382 rule implements the notion of two people knowing one another in terms of
383 their group membership - i.e. they know one another if they are from the

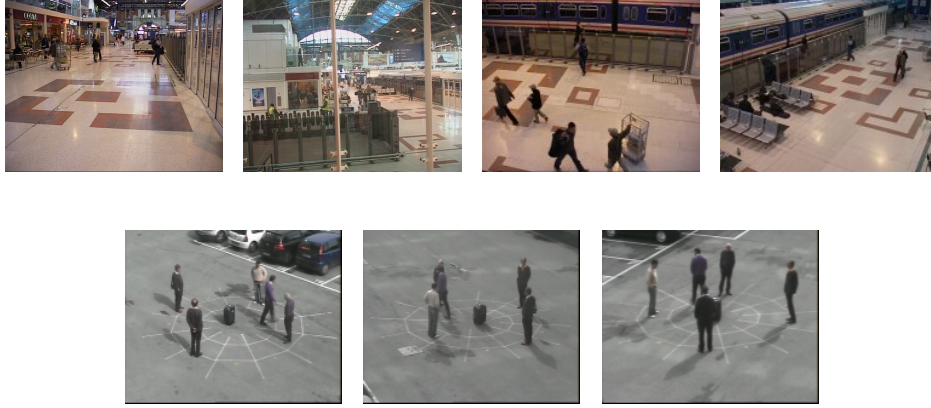


Figure 5: Datasets used. Top row: Four views from PETS2006 which contains scenarios with abandoned luggage. Bottom row: Three views from the SUBITO dataset describes scenarios where luggage owner enters the scene, sometimes interacts with other individuals and leaves the scene with/without the luggage.

408 doned bag scenario at a train station. All four camera views in the dataset
 409 were used in turn for the first four sequences used (PETS-S1-1, PETS-
 410 S1-2, PETS-S1-3 and PETS-S1-4), and camera view 3 used only for the
 411 other sequences (PETS-S2-3, PETS-S3-3, PETS-S4-3, PETS-S5-3, PETS-
 412 S6-3 and PETS-S7-3). The SUBITO dataset was recorded specifically for
 413 the SUBITO project. It contains thirteen sequences (19-22, 24-29, 31, 36,
 414 37) each recorded from four synchronised cameras placed around the scene.
 415 In sequences 19-22 a single person brings a bag to a marked position and lo-
 416 ters around the bag (sequence 19), abandons the bag (sequence 20), or leaves
 417 the bag unattended for a while and then comes back (sequences 21, 22). Se-
 418 quences 24-29, 31, 36 and 37 contain more challenging variants in terms of
 419 number of people and the group relationships. Each action is recorded 12

Table 1: Aggregate results across all SUBITO sequences comparing predicted alarms with corresponding baseline/owner/group ground truth.

Ruleset	TP	GTalarms	Alarms	Recall	Precision
baseline	16	71	35	0.23	0.46
owner	48	143	75	0.34	0.64
group	39	107	66	0.36	0.59

times for different entrance/exit directions. Depending on different threat definitions, the same action may or may not raise an alarm. Each sequence therefore should either correspond to 12 alarms (except for sequence 36 which only corresponds to 11 alarms), or none. The ground-truth alarms were obtained manually for all three threat definitions. The alarm time is determined by first visually deciding the very frame when the owner is just outside the prescribed distance from the bag, then adding a fixed time interval before the alarm is raised. Within the SUBITO dataset, the critical distance around a bag is assumed to be 2.5 metres (as opposed to 2 metres used in the PETS2006 challenge)- this assumption is therefore used in the three threat definitions. The time a bag must remain unattended to raise an alarm is reduced to 4 seconds.

6.2. Preliminary experiments on PETS2006 data

In the first experiments, the baseline functionality of (PETS2006) was implemented and evaluated. These experiments were carried out using an earlier version of the threat assessment logic implemented in C++. This was subsequently re-implemented in Prolog as part of the real-time system. To achieve this, the Prolog is queried for an alarm on every frame, based on

Table 2: Aggregate results across all SUBITO sequences comparing the use of all three threat definitions with the ground truth for the *owner+group* definition.

Ruleset	TP	GTalarms	Alarms	Recall	Precision
baseline	16	107	35	0.15	0.46
owner	42	107	75	0.39	0.56
group	39	107	66	0.36	0.59

Table 3: Aggregate results across all SUBITO sequences comparing the use of all three threat definitions with the ground truth for the *owner+group* definition with *stitched-together tracks*.

Ruleset	TP	GTalarms	Alarms	Recall	Precision
baseline	15	107	36	0.14	0.42
owner	43	107	94	0.40	0.46
group	41	107	88	0.38	0.47

438 the current state of the world and pertinent facts from the recent past. This
 439 world model is continually refreshed with the current location of each tracked
 440 object.

441 For the threat assessment to be correct, the system is required to raise
 442 an alarm following a potential threat, and to correctly identify the ID of
 443 the abandoned bag. Specifically, an alarm must be raised within 50 frames
 444 of a ground-truth alarm for it to be successful detected. The results on
 445 the PETS2006 dataset employ automatic tracking using an implementation
 446 of Breitenstein et al. (2011) and bag detection using Porikli et al. (2008).
 447 Alarms were raised correctly on all tested sequences except PETS-S4-3 and

448 PETS-S7-3. The failures on these two sequences were caused by individu-
449 als, having nothing to do with the abandoned bag, nevertheless being close
450 enough to prevent the bag being classified as unattended. This result moti-
451 vates the concept of ownership considered in the main set of experiments.

452 6.3. Experiments on SUBITO data

453 The main set of experiments were carried out on the challenging SUBITO
454 dataset. The inverse SFM system is run in batch mode so that it has access
455 to an entire sequence in predicting social groups rather than only the history
456 up until the current time. The entire sequence is therefore used in inferring
457 the set of alarms. This enabled evaluation of the interaction of the detection
458 and tracking sub-system and the threat assessment sub-system, giving the
459 inverted SFM the best chance of assigning correct social groups within rela-
460 tively short scenarios. A single threshold in the inverse SFM system controls
461 the propensity of pairs of individuals to be combined into the same group;
462 a lower threshold results in larger social groups. For the SUBITO data, we
463 found that both precision and recall reach their highest values within a small
464 range of this threshold and the results we present are for a choice of threshold
465 in this range.

466 The aggregate results across all SUBITO sequences are shown in Table 1,
467 comparing predicted alarms with the corresponding ground-truth - that is
468 baseline results are compared with the baseline ground-truth, etc. The ag-
469 gregate results comparing the use of all three threat definitions with the
470 ground-truth for the *owner+group* definition are shown in Table 2. As ex-
471 pected, the precision and recall for the *baseline* definition are lower in this
472 case since the ground-truth reflects a more sophisticated notion of threat,

473 incorporating concepts that are not present in the *baseline* definition. The
474 evaluation reported here attended only to the time an alarm is raised and
475 ignored the ID for the person and bag involved. Where there is more than
476 one true positive alarm for a ground-truth alarm, this is counted once in com-
477 puting recall and does not contribute to loss of precision. In other words,
478 multiple predicted alarms for the same ground-truth alarm are counted only
479 once. In general, there were few instances of this occurring in the experiment.

480 Within Table 2, there is a clear improvement in precision and recall be-
481 tween *baseline* and *owner* definitions. However, the comparison of perfor-
482 mance between *owner* and *owner+group* definitions is less decisive. Here the
483 recall has reduced slightly with the introduction of the social relationships,
484 but there is a comparable improvement in precision. Looking in more detail
485 at the results on individual sequences and alarms, several alarms have been
486 suppressed by correct assignment of an owner and partner to the same social
487 group. This is illustrated in Figure 6 showing a set of frames from SUBITO
488 sequence 36. Two individuals (d:211, d:212) entering the scene (Figure 6
489 (top)) are assigned to the same social group (indicated by blue line between
490 them), and one is detected as the owner of a bag (d:212) that appears within
491 the scene (Figure 6 (middle)). The owner subsequently goes away from the
492 bag and outside the prescribed distance (shown as a green circle around the
493 bag), leaving their partner attending to the bag (Figure 6 (bottom)). No
494 alarm is raised.

495 In general the recall and precision are below acceptable performance for
496 a deployed threat assessment system. The principal source of error arises
497 from the highly challenging video sequences containing multiple overlapping

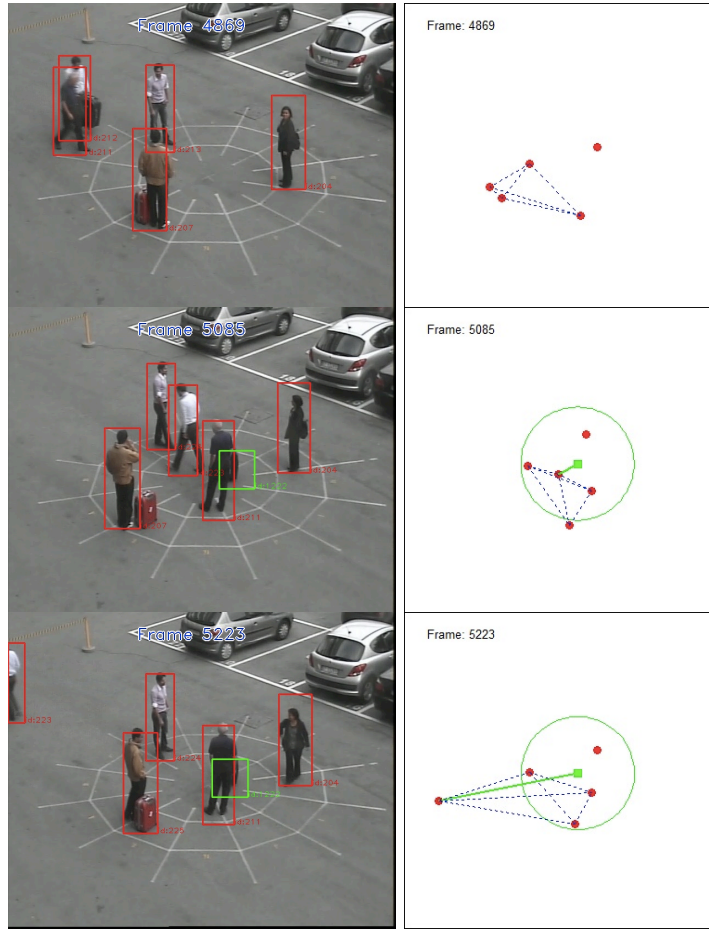


Figure 6: Social group analysis applied to SUBITO sequence 37 resulting in correct suppression of false alarm.

498 actors at any time. The consequential limitations in detection and tracking
 499 performance are translated directly into the threat assessments that can be
 500 achieved using the logic described above. Some improvement in performance
 501 was achieved by automatically stitching together tracks for which there is
 502 sufficient evidence that they belong to the same objects at different periods
 503 of time - specifically, one track (of more than 10 frames duration) ends within

504 4 seconds and 1 metre of another track (of more than 10 frames duration)
505 beginning. The precision and recall for the equivalent evaluation to that in
506 Table 2 is shown in Table 3. Finally, a real-time system that incorporates
507 all stages of the pipeline, including on-line estimation of social groups up to
508 the current frame, has also been implemented to demonstrate the practical
509 viability of the method.

510 7. Conclusions and Future Work

511 This paper has described a video surveillance framework that detects
512 abandoned objects in surveillance scenes containing multiple interacting in-
513 dividuals, extending the state of the art. Future work will address methods
514 to further improve the underpinning object (person and bag) detection and
515 tracking accuracy, as well as introduction of goal-directed and intentionality
516 modelling strategies in the behavioural analysis.

517 There is scope to perform a more rigorous analysis of ownership through
518 detecting bags being carried into the scene and hence identifying the owner
519 more reliably. Similarly, confidence that a bag has been removed from the
520 scene would be raised if it could be detected as it was carried out. There
521 is prior work on this problem that should in principle be directly applicable
522 to sequences such as those in the SUBITO dataset (e.g. Damen and Hogg
523 (2008)).

524 Finally, expressing the the conditions of a threat in terms of logic, sug-
525 gests that it may be possible to induce such conditions automatically from
526 examples, thereby providing a way to incorporate different kinds of informa-
527 tion about the scene without having to provide the logical rules by hand.

528 Earlier work on the use of inductive logic programming in video analysis
529 indicates how this might be achieved in principle (Dubba (2010)).

530 **Acknowledgements**

531 This work was supported by EC projects SUBITO Grant Agreement No.
532 218004 and FP7-ICT-247022 MASH. Any opinions expressed in this paper do
533 not necessarily reflect the views of the European Community. The Commu-
534 nity is not liable for any use that may be made of the information contained
535 herein.

536 H. Ardö and K. Aström., 2007. Multi Sensor Loitering Detection Using On-
537 line Viterbi. In Proc. IEEE Int. Workshop on Performance Evaluation of
538 Tracking and Surveillance (PETS), ISBN 0-7049-1423-9.

539 D. Arsić, M. Hofmann, B. Schuller and G. Rigoll., 2007. Multi-Camera Person
540 Tracking And Left Luggage Detection Applying Homographic Transforma-
541 tion. In Proc. IEEE Int. Workshop on Performance Evaluation of Tracking
542 and Surveillance (PETS), ISBN 0-7049-1423-9.

543 E. Auvinet, E. Grossmann C. Rougier, M. Dahmane and J. Meunier., 2006.
544 Left-Luggage Detection Using Homographies and Simple Heuristics. In
545 Proc. IEEE Int. Workshop on Performance Evaluation of Tracking and
546 Surveillance (PETS), ISBN 0-7049-1422-0.

547 A. F. Aveni., 1977. The Not-So-Lonely Crowd: Friendship Groups in Collec-
548 tive Behavior. Sociometry, 40(1):9699.

- 549 A. Bayona, J. C. SanMiguel, and J. M. Martínez., 2009. Comparative Eval-
550 uation of Stationary Foreground Object Detection Algorithms based on
551 Background Subtraction Techniques. In Proc. of the 6th IEEE Int. Con-
552 ference on Advanced Video and Signal Based Surveillance (AVSS 09),
553 DOI:10.1109/AVSS.2009.35, pp. 2530.
- 554 J. Berclaz, A. Shahrokni, F. Fleuret, J. Ferryman and P. Fua., 2009. Eval-
555 uation of Probabilistic Occupancy Map People Detection for Surveillance
556 Systems. In Proc. of the IEEE Int. Workshop on Performance Evaluation
557 of Tracking and Surveillance (PETS), pp 55-62, ISBN 978-07049-1501-4.
- 558 P. Birch, W. Hassan, N. Bangalore, R. Young and C. Chatwin., 2011. Sta-
559 tionary Traffic Monitor. In Proc. 4th Int. Conf. on Imaging for Crime
560 Detection and Prevention (ICDP-11), DOI:10.1049/ic.2011.0128 , pp. 1-6.
- 561 S. S. Blackman., 2004. Multiple Hypothesis Tracking for Multiple Target
562 Tracking. In IEEE Aerospace and Electronic Systems Magazine, 19(1),
563 pp. 5-18.
- 564 M. Brand, N. Oliver and A. Pentland., 1997. Coupled Hidden Markov Models
565 for Complex Action Recognition. In Proc. Computer Vision and Pattern
566 Recognition, pp. 994-999.
- 567 M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier and L. Van Gool.,
568 2011. Online Multiperson Tracking-by-Detection from a Single, Uncali-
569 brated Camera. In IEEE Trans. on Pattern Analysis and Machine Intelli-
570 gence, 33(9), pp. 1820-1833.

- 571 R. Challenger, C. W. Clegg, M. A. Robinson, and M. Leigh., 2009. Under-
572 standing Crowd Behaviours: Simulation Tools. Technical Report, Univer-
573 sity of Leeds.
- 574 G. Dalley, X. Wang, and W.E.L. Grimson., 2007. Event Detection Using an
575 Attention-Based Tracker. In Proc. IEEE Int. Workshop on Performance
576 Evaluation of Tracking and Surveillance (PETS), ISBN 0-7049-1423-9.
- 577 D. Damen and D. Hogg., 2008. Detecting Carried Bags in Short Video Se-
578 quences. In Proc. 10th European Conf. on Computer Vision, 5304:154-167.
- 579 D. Damen and D. Hogg., 2012. Explaining Activities as Consistent Groups
580 of Events: A Bayesian Framework Using Attribute Multiset Grammars. In
581 Int. Journal of Computer Vision, 98(1), pp. 83-102, DOI:10.1007/s11263-
582 011-0497-0.
- 583 K. S. R. Dubba, A. G. Cohn, and D. C. Hogg., 2010. Even Model Learning
584 From Complex Videos using ILP. In Proceedings ECAI, 215, pp. 93-98.
- 585 R. Evangelio and T. Sikora., 2011. Static Object Detection Based on a Dual
586 Background Model and a Finite-State Machine. EURASIP Journal on Im-
587 age and Video Processing, Article ID 858502, DOI:10.1155/2011/858502.
- 588 W. Ge, R. Collins, and B. Ruback., 2009. Automatically Detecting the Small
589 Group Structure of a Crowd. In Workshop on Applications of Computer
590 Vision, pp. 18.
- 591 S. Guler and M. K. Farrow., 2006. Abandoned Object Detection in Crowded
592 Places. In Proc. IEEE Int. Workshop on Performance Evaluation of Track-
593 ing and Surveillance (PETS), ISBN 0-7049-1422-0.

594 D. Helbing and P. Molnar., 1995. Social Force Model for Pedestrian Dynam-
595 ics. Physical Review E, 51:4282.

596 S. Hongeng, R. Nevatia and F. Brémond., 2004. Video-Based Event Recog-
597 nition: Activity Representation and Probabilistic Recognition Methods.
598 Computer Vision and Image Understanding, 96:129-162.

599 Imagery Library for Intelligent Detection Systems, *http : //www.ilids.co.uk*.
600 Last accessed: 29 January 2012.

601 Y. Ivanov and A. Bobick., 2000. Recognition of Visual Activities and Inter-
602 actions by Stochastic Parsing. In IEEE Trans. on Pattern Analysis and
603 Machine Intelligence, 22(8), pp. 852-872.

604 J. Jacques, A. Braun, J. Soldera, S. Musse, and C. Jung., 2007. Understand-
605 ing People Motion in Video Sequences using Voronoi Diagrams. Pattern
606 Analysis and Applications, 10:321332.

607 S-W. Joo and R. Chellappa., 2006. Attribute Grammar-Based Event Recog-
608 nition and Anomaly Detection. Proceedings Int. Workshop on Semantic
609 Learning Applications in Multimedia, New York , NY June.

610 N. Krahnstoever, P. Tu, T. Sebastian, A. Perera and R. Collins., 2006. Multi-
611 View Detection and Tracking of Travelers and Luggage in Mass Transit
612 Environments. In Proc. IEEE Int. Workshop on Performance Evaluation
613 of Tracking and Surveillance (PETS), ISBN 0-7049-1422-0.

614 V. Leung and S. Herbin., 2011. Flexible Tracklet Association for Complex
615 Scenarios using a Markov Logic Network. In Proc. 11th Int. Workshop on
616 Visual Surveillance, pp. 1870-1875.

- 617 L. Li, R. Luo, R. Ma, W. Huang and K. Leman., 2006. Evaluation of an
618 IVS System for Abandoned Object Detection on PETS 2006 Datasets. In
619 Proc. IEEE Int. Workshop on Performance Evaluation of Tracking and
620 Surveillance (PETS), ISBN 0-7049-1422-0.
- 621 J. Li, S. Gong and T. Xiang., 2008. Global Behaviour Inference using Prob-
622 abilistic Latent Semantic Analysis. In Proc. British Machine Vision Conf.,
623 DOI:10.5244/C.22.20.
- 624 S. Lu, J. Zhang and D. Feng., 2007. Detecting Unattended Packages through
625 Human Activity Recognition and Object Association. Pattern Recognition,
626 8:2173-2184.
- 627 F. Lv, X. Song, B. Wu, V. K. Singh and R. Nevatia., 2006. Left-Luggage
628 Detection using Bayesian Inference. In Proc. IEEE Int. Workshop on Per-
629 formance Evaluation of Tracking and Surveillance (PETS), ISBN 0-7049-
630 1422-0.
- 631 J. Martínez-del-Rincón, J. Herrero-Jaraba, J. Ral Gómez and C. Orrite-
632 Uruñuela., 2006. Automatic Left Luggage Detection and Tracking Using
633 Multi-Camera UKF. Proceedings IEEE Int. Workshop on Performance
634 Evaluation of Tracking and Surveillance (PETS), ISBN 0-7049-1422-0.
- 635 M. Moussaid, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G.
636 Theraulaz., 2009. Experimental Study of The Behavioural Mechanisms
637 Underlying Self-Organization in Human Crowds. In Proc. of the Royal
638 Society B: Biological Sciences, 276(1668), pp. 2755-2762.

639 M. Moussaid, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz., 2010.
640 The Walking Behaviour of Pedestrian Social Groups and Its Impact on
641 Crowd Dynamics. PLoS ONE, 5(4):e10047, 04.

642 S. Pellegrini, A. Ess, and L. Van Gool., 2010. Improving Data Association
643 by Joint Modeling of Pedestrian Trajectories and Groupings. In Proc. Eu-
644 ropean Conference on Computer Vision. 6311:452465, Springer Berlin /
645 Heidelberg.

646 PETS2006. 2006. Ninth IEEE Int. Workshop on Performance Evaluation of
647 Tracking and Surveillance (PETS), New York, USA, June 18, ISBN 0-
648 7049-1422-0, ISBN 0-7049-1422-0.

649 PETS2007. 2007. Tenth IEEE Int. Workshop on Performance Evaluation of
650 Tracking and Surveillance (PETS), Rio De Janeiro, October 14, ISBN 0-
651 7049-1423-9.

652 F. Porikli and Z. Yin., 2006. Temporally Static Region Detection in Multi-
653 Camera Systems. In Proc. IEEE Int. Workshop on Performance Evaluation
654 of Tracking and Surveillance (PETS), ISBN 0-7049-1422-0, ISBN 0-7049-
655 1422-0.

656 F. Porikli, Y. Ivanov and T. Haga., 2008. Robust Abandoned Object De-
657 tection using Dual Foregrounds, EURASIP Journal on Advances in Signal
658 Processing, Article ID 197875.

659 C. W. Reynolds. 1987. Flocks, Herds and Schools: A Distributed Behavioral
660 Model. In Proc. of the 14th Annual Conference on Computer Graphics and
661 Interactive Techniques, pp. 2534, DOI:10.1145/37401.37406.

- 662 P. C. Ribeiro, P. Moreno and J. Santos-Victor., 2007. Detecting Luggage
663 Related Behaviors Using a New Temporal Boost Algorithm. In Proc. IEEE
664 Int. Workshop on Performance Evaluation of Tracking and Surveillance
665 (PETS), ISBN 0-7049-1423-9.
- 666 M. Richardson and P. Domingos., 2006. Markov Logic Networks, Machine
667 Learning, 62:107136.
- 668 C. Rudloff, T. Matyus, and S. See and D. Bauer., 2011. Can Walking Be-
669 haviour Be Predicted? An Analysis of the Calibration and Fit of Pedes-
670 trian Models. 90th Annual Meeting of the Transportation Research Board,
671 January.
- 672 V. D. Shet, D. Harwood and L. S. David., 2005. Vidmap: Video Monitoring
673 of Activity with Prolog. In Proc. IEEE Conference on Advance Video and
674 Signal Based Surveillance, pp. 224-229.
- 675 Y. Shi, Y. Huang, D. Minnen, A. Bobick and I. Essa., 2004. Propagation
676 Networks for Recognition of Partially Ordered Sequential Action. In Proc.
677 IEEE Computer Society Conference on Computer Vision and Pattern
678 Recognition, 2:862-869, DOI:10.1109/CVPR.2004.1315255.
- 679 H. Singh, R. Arter, L. Dodd, P. Langston, E. Lester and J. Drury., 2009.
680 Modelling Subgroup Behaviour in Crowd Dynamics Dem Simulation. Ap-
681 plied Mathematical Modelling, 33(12):44084423.
- 682 K. Smith, P. Quelhas and D. Gatica-Perez., 2006. Detecting Abandoned Lug-
683 gage Items in a Public Space. In Proc. IEEE Int. Workshop on Performance
684 Evaluation of Tracking and Surveillance (PETS), ISBN 0-7049-1422-0.

- 685 J. Sochman and D. C. Hogg., 2011. Who Knows Who - Inverting the Social
686 Force Model for Finding Groups. In Proc. IEEE Intelligent Workshop on
687 Socially Intelligent Surveillance and Monitoring.
- 688 Y. Tian, R. Feris, H. Liu, A. Humpapur, and M-T. Sun., 2010. Robust
689 Detection of Abandoned and Removed Objects in Complex Surveillance
690 Videos, In Proc. IEEE Trans. on Systems, Man, and Cybernetics, Part C:
691 Applications and Reviews, pp. 112.
- 692 S. Tran and L. Davis., 2008. Visual Event Modelling and Recognition Using
693 Markov Logic Networks. In Proc. European Conf. on Computer Vision,
694 pp. 610-623.
- 695 X. Wang, X. Ma and E. Grimson., 2009. Unsupervised Activity Perception
696 in Crowded and Complicated Scenes Using Hierarchical Bayesian Models.
697 IEEE Trans. on Pattern Analysis and Machine Intelligence, 31(3):539-555.
- 698 A. Yildiz and Y. S. Akgul., 2010. A Fast Method for Tracking People with
699 Multiple Cameras, Technical Report, Vision Lab, Gebze Institute of Tech-
700 nology.
- 701 Z. Zivkovic. 2004. Improved Adaptive Gaussian Mixture Model for Back-
702 ground Subtraction, In Proc. Int. Conf. on Pattern Recognition, 2, pp.
703 28-31.